LEUPHANA
UNIVERSITÄT LÜNEBURG

LEARNING FROM DATA

01. Introduction

Who am I?

Prof. Dr. Burkhardt Funk

— Professor for Data Science at Leuphana (Institute of Information Systems)
— Program Coordinator Major Business Information Systems (College)
— Several administrative functions (Promotionskommission, FKR, Prüfungsausschuss Bachelor/Master)

Short CV

— Married, 2 daughters, sailor
— Professor for IS and Data Science – initially the driving force behind the Data Science Program, visiting scholar at University of Virginia, Stanford University
— 20 years of entrepreneurial experience (founded 10+ startups), advisor and board member → happy to discuss your entrepreneurial activities/ ideas
— PhD in computational physics (Wuppertal), studied physics & computer science in Kiel, Würzburg, Stony Brook/NY, Tsukuba/Japan

Research focus

Research contexts

- Health
- Information Extraction
- E-Commerce
- Sensors and IoT

Methods

- Machine learning
- Computer vision
- Bayesian statistics

Projects

GET.ON Gesundheits Training.Online

Digi-Exist

reconnected

E-COMPARED

ZIL

HBI

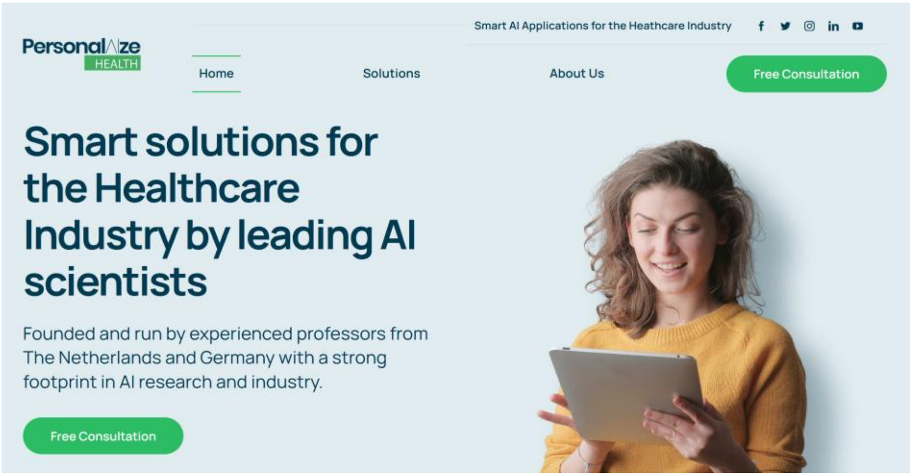
CARE

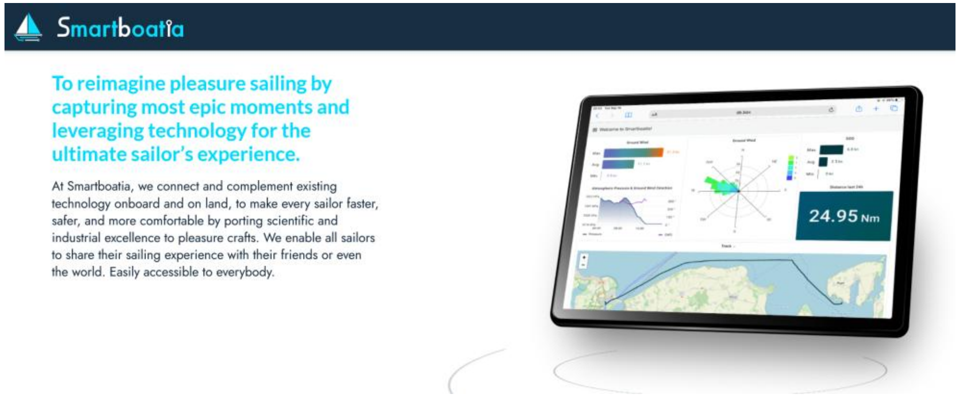
SEM-A²

DATAx

https://scholar.google.de/citations?hl=de&user=Ag3XW6wAAAAJ&view_op=list_works&sortby=pubdate

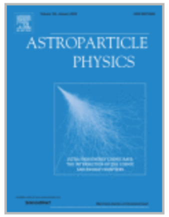
Startup activities





... having a long relationship with AI

Astroparticle Physics
Volume 4, Issue 2, December 1995, Pages 119-132



Separating γ- and hadron-induced cosmic ray air showers with feed-forward neural networks using the charged particle information ⋆

S. Westerhoff [a], B. Funk [a], A. Lindner [b], N. Magnussen [a], H. Meyer [a], H. Möller [a], W. Rhode [a], R.N. Sooth [a], B. Wiebel-Sooth [a]

Teaching team

Exercise:

— Jonas Scharfenberger
— PhD student (does fancy stuff in computer vision – ask him)
— M.Sc. in Mathematics (WWU Münster)

Tutorial:

— Moritz Burmester
— Pursuing the M.Sc. Management and Data Science program at Leuphana (5th Semester)

Introduce yourself

- Write a short CV (1-Pager, including your photo, academic background, data science experience, fun fact)
- Upload to *mystudy>LFD>profile* until **Friday, Oct. 17th**

Want to work with us as a student assistent?

All first-year students are required to enroll in a course focused on programming and data analysis using Python. As part of the coursework, one of the assignments is to develop a simple program, provide a detailed description of it, and create a short video presentation. We aim to provide automatic feedback using LLM-based technology...

- Mini Coding Projekt
- Empfohlene Aufgaben für ZuhauseA...
- 4. Ergänzende MaterialienDie Welt de...
- Ergänzende Videos ansehen Zum Th...
- 5. Geben Sie uns FeedbackUnten hab...
- Wir brauchen Ihr Feedback, um bess...
- Kapitel 6: Datenvisualisierungs...
- 1. Einleitung (Kopie) (Kopie) (Kopie) ...
- Zur Einordnung auf diese WocheDies...

Enchanted Forest of Fairypine:

For my Mini Coding Project, i programmed a little interactive text-based adventure that leads the player into a magical world filled with mystical creatures. The player journeys through various sections of the forest and has to make decisions that will influence the course of their adventure.

If the player decides to play the game, they find themselves in a forest that is filled with magical creatures. The goal of the game is to find the heart of the forest. The player navigates the game by making decisions, at different crossroads and in various situations. They can choose from different paths that lead to different encounters and events. The game ends either when the player reaches the heart of the forest and therefore wins the game or by failing.

I am aware that the game is not yet perfect, but i wanted to try out different ways of programming an interactive game and learn how to programme linked pathways. However, in my journey of programming i did have some help from Chat GPT who showed me how a construct like that is programmed in general. I learned how to Programme an interactive story-game. I learned how to add new paths and tasks, more detail and context, but also possibilities of linking different paths with different choices the player makes. At the moment the game is very simple, since it does take a lot of time to create a whole story.

Anyways, Have fun!

Today's objectives

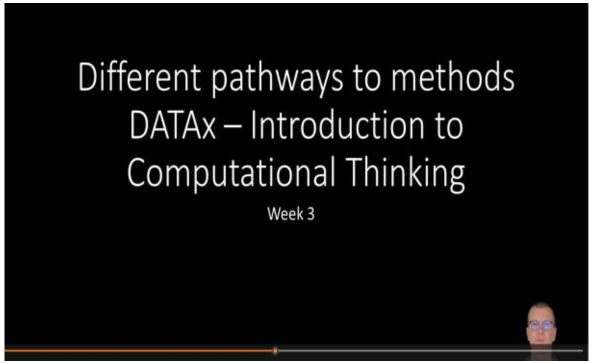
- Overview and introduction of the course
- Administrative and organizational topics
- Get an intuition of what machine learning means
- Define a learning problem
- Intro to the mother of all learning algorithms

This course is not about the fancy AI stuff …

Multiagent Frameworks

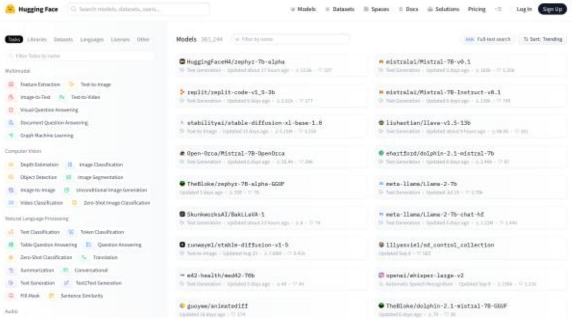
https://www.youtube.com/watch?v=Zlgkzjndpak

On-the-fly translation



Style transfer

Explore it (huggingface)



… but the conceptual underpinning of machine learning

Learning from Data

=

Conceptual Foundation

+

Techniques

Agenda

- **Introduction**
- **Learning Problem**
- Linear Models (Regression)
- Linear Models (Classification)
- Decision Trees
- Learnability, Error, and Noise
- Theory of Generalization
- Bias-Variance Tradeoff
- Support Vector Machines
- Neural Networks
- Non-linear Transformations
- Overfitting and Regularization
- Ensemble Methods
- Unsupervised Learning

Admin and organization

— Lecture: Tuesday 10:15
— Help us to make the lecture an interactive experience (ask questions early on, rephrase insights, prepare the lecture)
— Exercises/Lab: Thursday 12:15
— **Tutorial: Thursday 16:15**
— Weekly assignments: be prepared; for implementation tasks we will use Python
— Written exam: the thing that counts, but hardly helps to become an expert (preliminary dates: to be done)

Resources – selected textbooks

- **Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H. T. (2012).** ***Learning from data*** **(Vol. 4). New York, NY, USA:: AMLBook.**
- Friedman, J., Hastie, T., & Tibshirani, R. (2013). *The elements of statistical learning*. New York: Springer series in statistics.
- S. Shalev-Shwartz & S. Ben-David (2014). Understanding Machine Learning. Cambridge
- Tom Mitchell (1997) Machine Learning. McGraw-Hill
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Kevin P. Murphy (2012) Machine Learning – A Probabilistic Perspective. MIT Press
- ***Additional material will be provided via mystudy***

Other freely accessible online resources

- **Yaser Abu-Mostafa – Learning from Data. (https://www.youtube.com/watch?v=mbyG85GZ0PI&list=PLD63A284B7615313A)**
- Andrew Ng – Machine Learning. (https://www.youtube.com/watch?v=PPLop4L2eGk&list=PLLssT5z_DsK-h9vYZkQkYNWcItqhlRJLN)
- Nando de Freitas – Machine Learning. (https://www.youtube.com/watch?v=w2OtwL5T1ow&list=PLE6Wd9FR--EdyJ5lbFl8UuGjecvVw66F6&index=1)
- Towards Data Science blog on Medium (https://towardsdatascience.com/machine-learning/home)

How it all started – the Dartmouth proposal

6. Organization expenses of $200. (Includes expense of reproducing preliminary work by participants and travel necessary for organization purposes.

7. Expenses for two or three people visiting for a short time.

<u>Estimated Expenses</u>

6 salaries of 1200	$7200
2 salaries of 700	1400
8 traveling and rent expenses averaging 300	2400
Secretarial and organizational expense	850
Additional traveling expenses	600
Contingencies	550
	$13,500

http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf

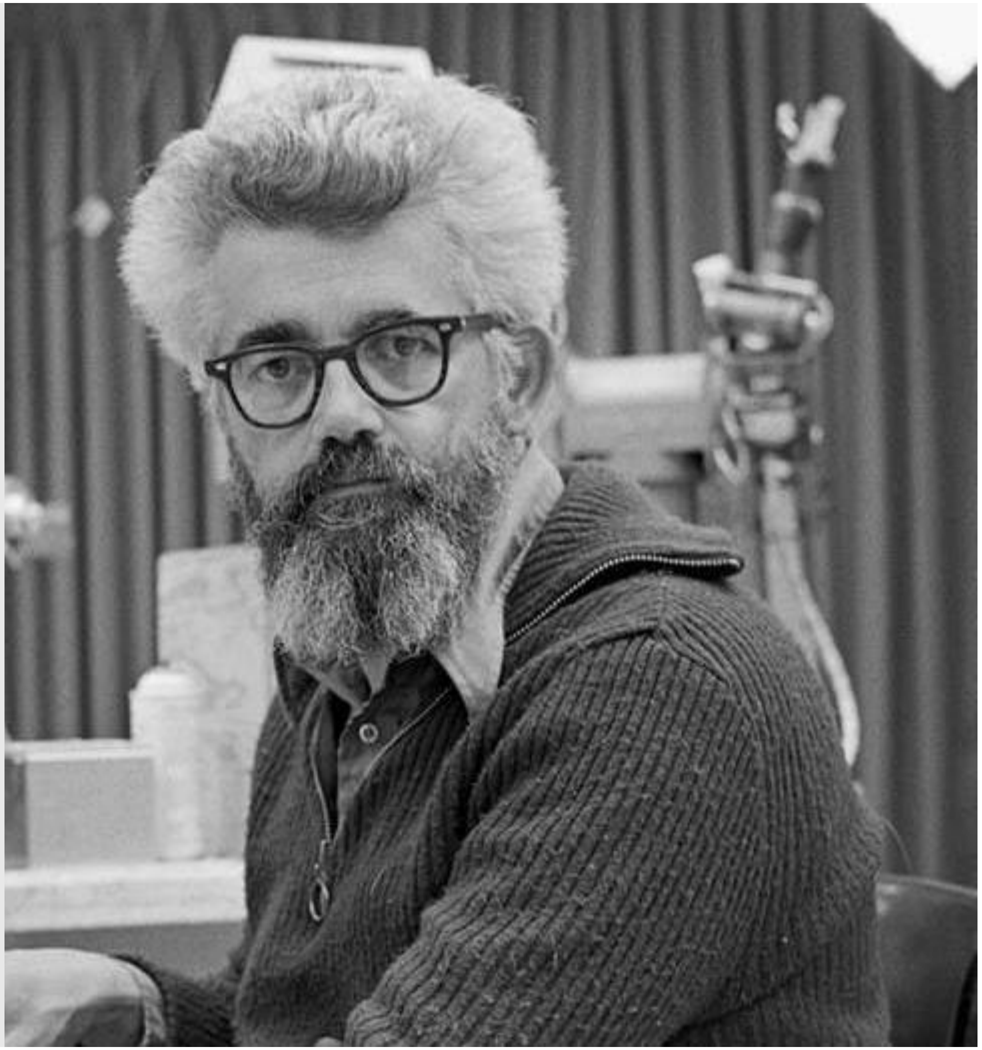
Two guys – a few years too late in Dartmouth

Artificial Intelligence

Artificial Intelligence (AI) tries to make computers and things smart

AI = “human intelligence exhibited by machines”

Examples: rule-based and expert systems, knowledge representation & reasoning, machine learning



John McCarthy

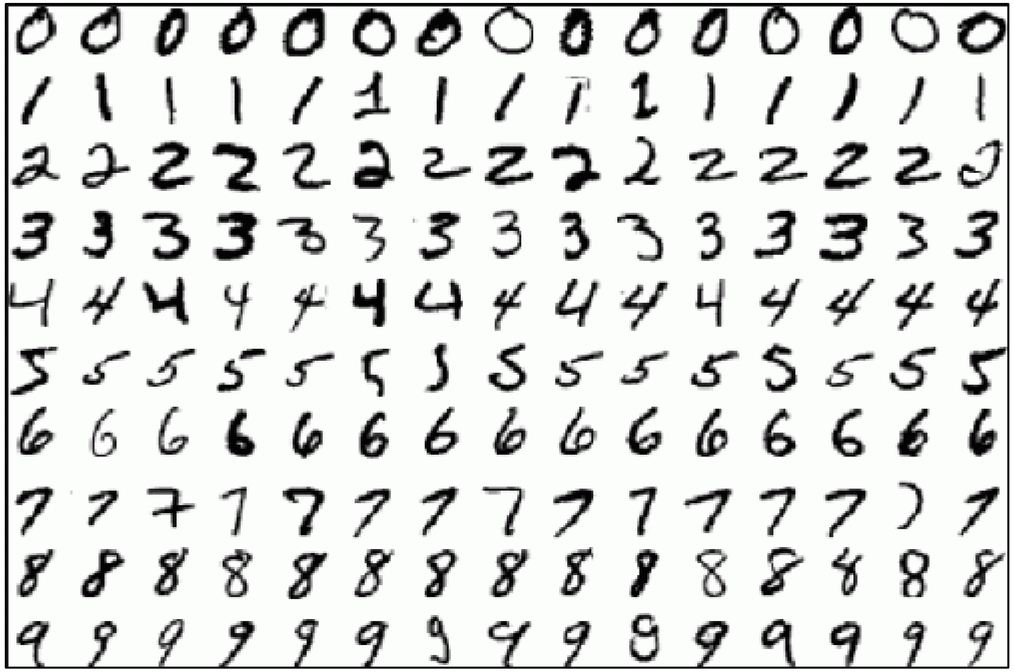
Machine learning is a subset of AI

"You can think of machine learning and artificial intelligence as a set of Russian dolls nested within each other. Machine learning is a subset of AI, which is an umbrella term for any computer program that does something smart."

Source: skymind.ai

The mother of all ML examples

- **MNIST dataset**; contains 60,000 handwritten digits with labels (28x28 pixel images)
- Algorithms try to recognize the digits by looking at the pixel details
- Many implementations for recognition exist; the best algorithms reach an error rate of below 1%



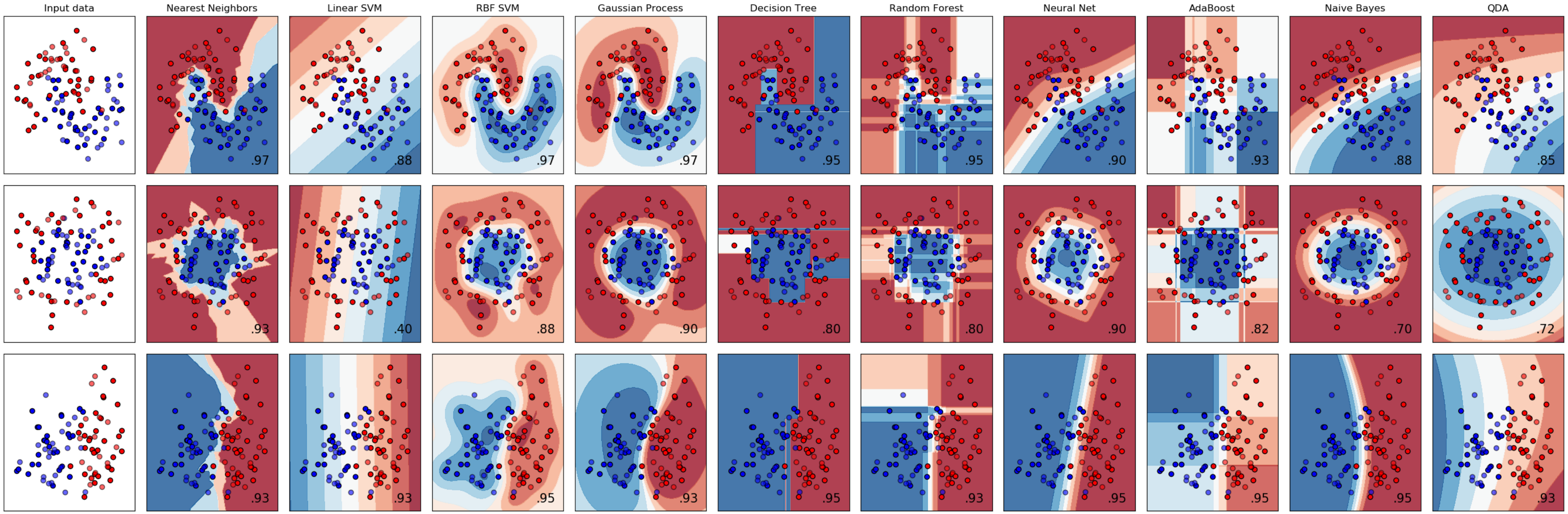
Source: Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, november 1998

Time for a quiz – MNIST dataset

1. What makes it hard for a computer to learn the task of recognizing handwritten digits?

2. What handcrafted **features** (properties of the images) could you think of to label handwritten digits automatically?

Today, fitting and applying simple ML models is „easy“



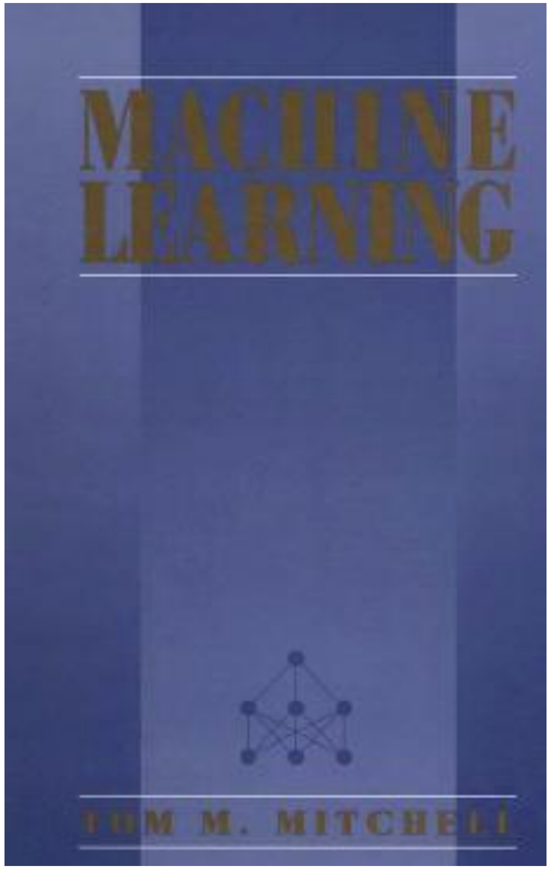
https://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html#sphx-glr-auto-examples-classification-plot-classifier-comparison-py

Defining a learning problem

Definition: A computer program is said to **learn** from experience **E** with respect to some class of tasks **T** and performance **P**, if its performance at tasks in T improves with E

Example: Learn to play checkers

- T: play checkers
- P: % of games won
- E: opportunity to play against itself

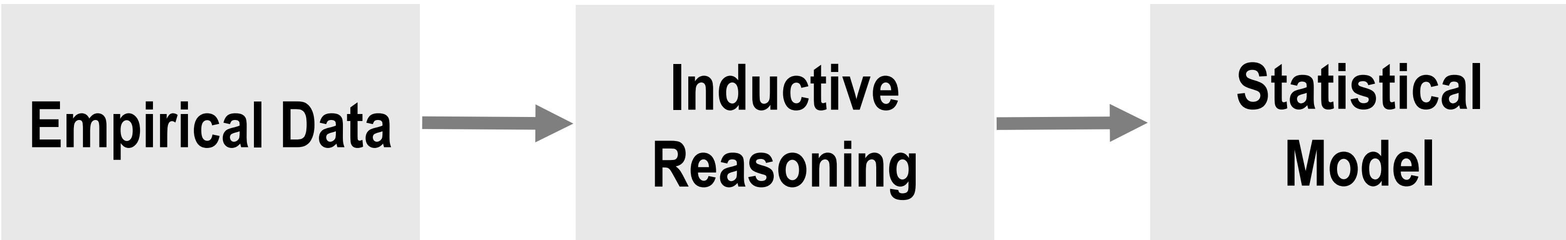


Tom Mitchell (1998) Machine Learning

Machine learning

— Machine learning is an „approach to achieve AI through systems that can learn from experience (data) to find patterns“

— We try to teach a computer to recognize patterns by providing examples, **rather than programming** it with specific rules.



Source: Jason Mayes (2017)

Types of learning

We have different types of learning, here are 4 important settings that you find in practice

- **Supervised learning (input, labels)**
- Unsupervised learning (input)
- Semi-supervised learning (lots of input, some observations with labels)
- Reinforcement learning

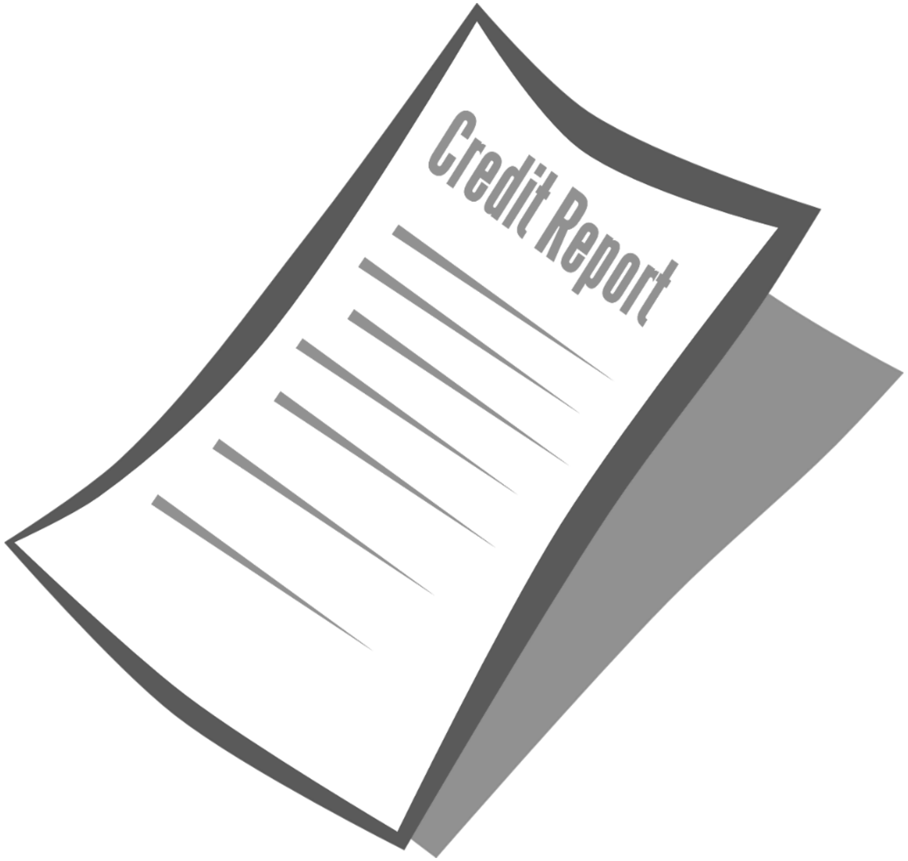
Notation and terminology

Key issues in machine learning

Transform real-world problem into computable learning problem

— **Define label space**: what do we want to learn (e.g. playing checkers: board move vs. board value)?
— **Determine type of training data**/ experience (e.g. given/biased, self-selected?)
— **Represent input data** (feature definition and extraction, functional form)
— Learning algorithm
— Evaluation (Is it a good model? Is it good enough?)
— What decisions shall be informed?

Credit scoring – a learning example



Source: https://www.needpix.com/

- Suppose we want to inform credit decisions of a bank. What information is included in a credit request from a customer?
- What features will have a positive impact on the credit decision?
- How could the set of hypothesis be represented?

A simple decision model

— Input vector $\boldsymbol{x} = [x_1, \dots, x_d]^T$

— Find weights for different inputs and compute *credit score*

$$\textit{credit score} = \sum_{i=1}^{d} w_i x_i \,.$$

— Approve credit if the *credit score* is acceptable.

Approve credit if $\sum_{i=1}^{d} w_i x_i \geq$ threshold, (*credit score* is good)

Deny credit if $\sum_{i=1}^{d} w_i x_i <$ threshold. (*credit score* is bad)

— **How to choose the "importance" weights w_i**

input x_i is important ⇒ large weight $|w_i|$

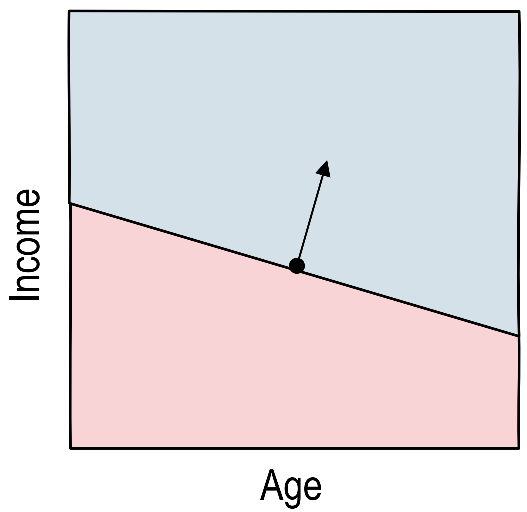
input x_i beneficial for credit ⇒ positive weight $w_i > 0$

input x_i detrimental for credit ⇒ negative weight $w_i < 0$

Simple decision model – revised version

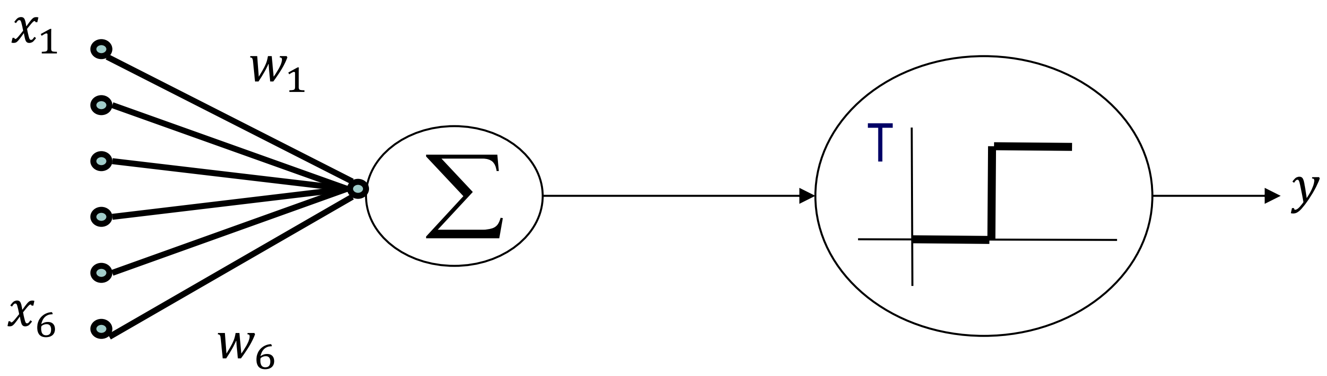
Visualizing the decision boundary

$$h(\mathbf{x}) = \mathrm{sign}(\mathrm{w}^{\mathrm{T}}\mathrm{x})$$



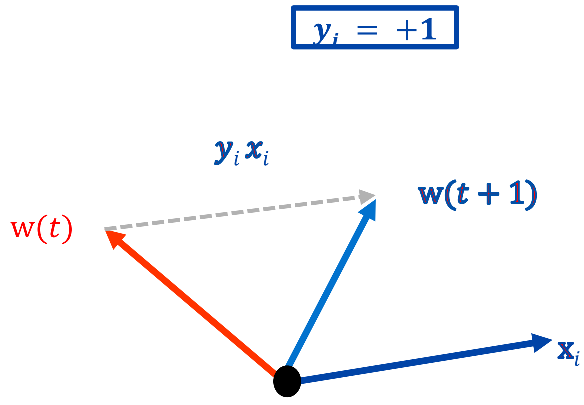
Online Learning - Perceptron

- Rosenblatt (1959) suggested that when a target output value is provided for a single neuron with fixed input, it can incrementally change weights and learn to produce the output using the Perceptron learning rule
- Online learning ("as new data comes in"), mistake driven algorithm
- Perceptron = Linear Threshold Unit



Source: Dan Ross: CS 446: Machine Learning

Perceptron Algorithm – basic idea



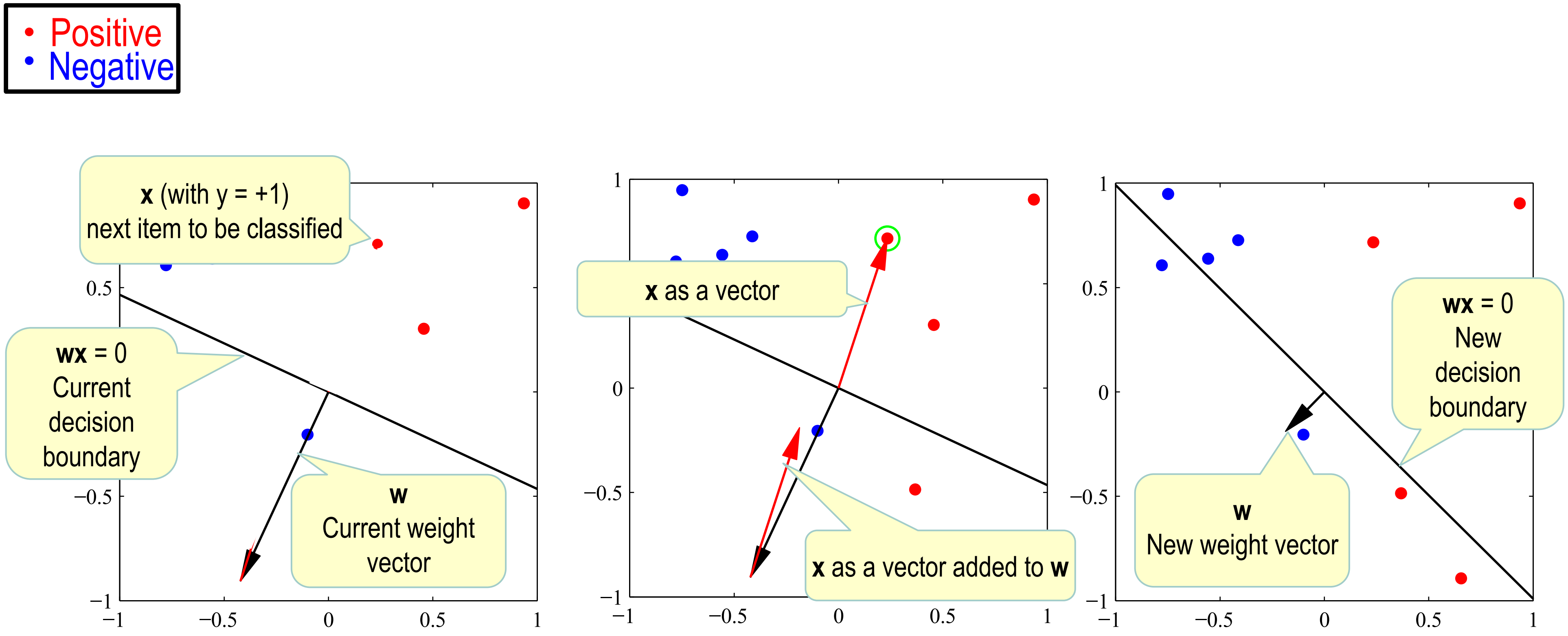
Assume that data is linearly separable

```
Start with arbitrary w(1) = 0
Do until all training data is correctly classified
      Pick any misclassified example (x_i, y_i)
      Update the weight:
```

$$\mathrm{w}(t+1) = \mathrm{w}(t) + \eta y_i x_i$$

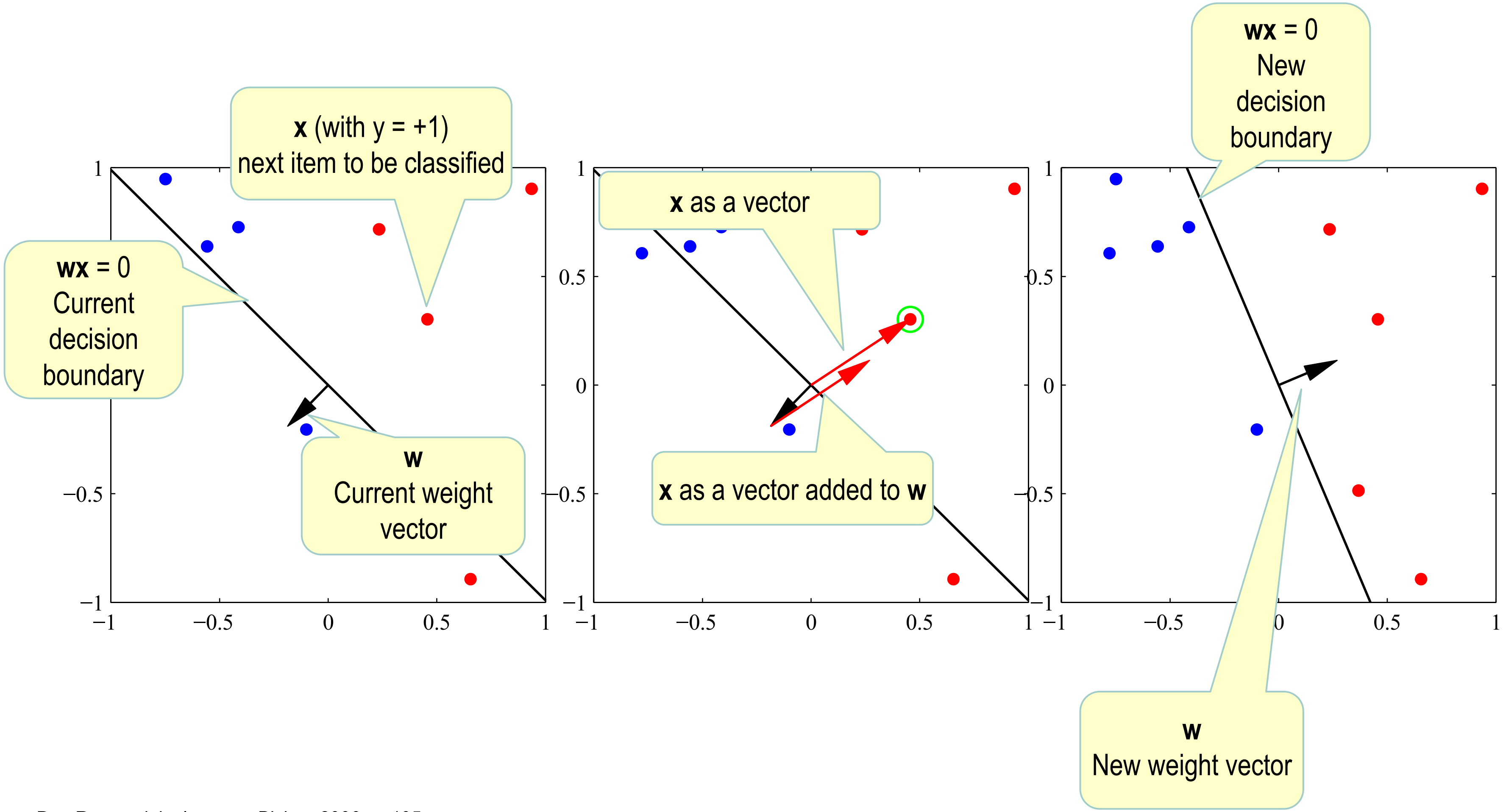
PLA implements incremental learning, how should we choose the learning rate?

How the algorithm adjusts the weight vector



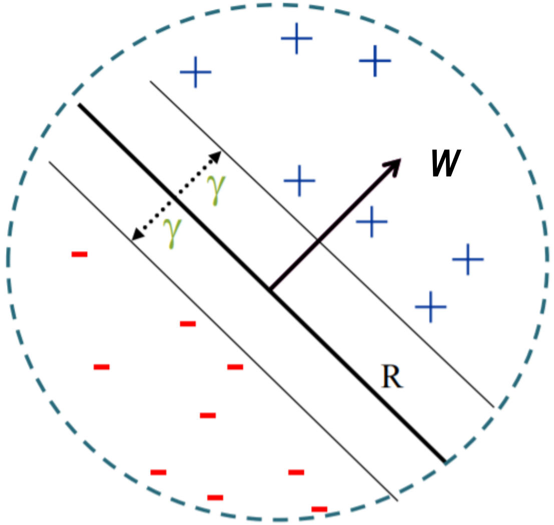
Source: Dan Ross, original source: Bishop 2006, p. 195

How the algorithm adjusts the weight vector



Source: Dan Ross, original source: Bishop 2006, p. 195

Number of mistakes the PLA makes is bound



Theorem (Novikoff 1962)

Let S be a training set and let

$$R = \max_{1 \leq i \leq n} ||x_i||_2$$

Suppose that there exists a vector w such that $||w|| = 1$ and $\gamma > 0$

$$y_i w^T x_i > \gamma$$ => *data is linearly separable*

for $1 \leq i \leq n$. Then the number of mistakes made by the online perceptron algorithm on S is at most

$$\left(\frac{R}{\gamma}\right)^2$$

Source: Machine Learning Department, School of Computer Science, Carnegie Mellon University

For a proof see e.g. Andrew Ng lecture notes (lecture 6 large margin classifiers)

Concluding remarks

— PLA always converges if data is linearly separable

— PLA and problem setting have everything we discussed to define a learning problem

— Try applying this to the Australia Rain dataset to predict the occurrence of rainfall. Does it converge? Why not?

Solution: Multi-Layer Perceptron (Neural Networks) or feature transformation

Historical remarks:

— Frank Rosenblatt simulated PLA on an IBM 704 in 1957 – built special hardware, were able to recognize characters from “photos”

— Already simple mechanisms can generate data that are not linearly separable (e.g. XOR) as discussed by Minsky and Papert (1969) → lead to a stop of research on neural networks